

A COMPUTATIONAL ANALYSIS OF ACHILLEA MILLEFOLIUM SEQUENCE TO DETERMINE EXPECTATION VALUE COMMONLY USED IN BIOINFORMATICS DATABASE

UMA KUMARI

Assistant Professor, Department of Biotechnology, Marwari College, Ranchi University, Ranchi, India

ABSTRACT

Achillea Millefolium economically important crop worldwide, intensively investigated and model system for genetic studies in plant. *Achillea L.* (Compositae or Asteraceae) distributed widely as a medicinal plant throughout the world and has been used since ancient time as a variability is a measure spread of data set. Genome analysis and annotation using genome from the libraries, automatic annotation using the BLAST (basic local alignment search tools) low complexity sequence has unusual composition that can create a problem in sequence similarity searching of *Achillea* protein sequence the color bars in the graphic summarize the BLAST. Database such as NCBI, FASTA, BLAST, ClustalW. BLAST has a bioinformatics tool that act as an algorithm to align sequences as if they were found in the database search. When expect value is increases from default value, a larger list with more scoring hits alignment can be reported. The desired subset of the plant *Achillea* data compiled using BLAST tool can be subsequently used for check the expectation value to analyses and knowledge discovery as the observation.

KEYWORDS: Biological Database, Customized Data Retrieved, Sequence Analysis, Data Compiled, Expectation Value *Achillea* & Asteraceae

Received: Dec 21, 2018; **Accepted:** Jan 11, 2018; **Published:** Jan 24, 2019; **Paper Id.:** IJBTRJUN20191

INTRODUCTION

The data of *Achillea* data searching allows the user to select one sequence and perform pairwise sequence interaction, pathway and system data. Pairwise alignment way to identify related sequence along with a scoring scheme, protein sequences or nucleotide sequence (Higgins Des, Taylor Willie2000). In *Achillea* contain primary data give the annotation and interpretation of data with the help of bioinformatics tools and server are freely available. NCBI E-utilities provide customized data utilities for various databases available at NCBI. These utilities require generation of URLs (Utility web services (<http://www.ncbi.nlm.nih.gov>) can be generated for the user in a specific for respective database either by manually or by writing scripts. The branch of computational biology finds out the sequence similarity search between the sequences. *A.Millefolium* use as a medicinal plant used in folk medicine to treat in many diseases also like inflammation, pain and gastrointestinal disorders. A sequence in FASTA format is represented as a series of lines, each of which should no longer than 120 character and usually do not exceed 80 character. Although, multiple URL can be generated. As the initial step of reach these goals, Casey, R.M., (2005).

MATERIALS AND METHODS

Genome sequence analysis of the yarrow plant by applying the sequence alignment method in computer programming for bioinformatics database. National center of biotechnology information focuses on theoretical, analytical, statistical and applied computational approaches and widely used primary database such as European nucleotide archive, Uniprot kb/Swiss –prot, a widely used method for assignment of secondary structure, the sequences in Operating system—UNIX.

Database and Corresponding Web Services

Database name	Web services type:	URL
NCBI E	Utility web services	(http://www.ncbi.nlm.nih.gov)
BLAST	www.ebi.ac.uk/tools/sss/NCBIBLAST	
FASTA	www.ebi.ac.uk/tools	
EMBL/EBI	EMBL-EBI web services	(http://www.ebi.ac.uk/tools/)
Uniprot KB	Programmatic access services	(http://www.uniprot.org)

RESULTS AND DISCUSSIONS

The Query sequence of plant yarrow with colour key of alignment score, each bar represents the portion of another sequence that is not so much similar to the query sequence, red bar represents most similar sequence, pink bar represent less good. Searching the Database is observing score/ high score E-Value, genomic analysis sequence. The bioinformatics software support data from single entry as well as multiple entries. Biological data is exchanged among these databases on the daily basis and find the pair wise sequence alignment. NCBI is housed as a series of databases relevant to bioinformatics tools and services where biological data is available. Major sequence include gene bank for DNA sequences, protein sequence and pubmed. Epigenomic database of the NCBI (National center of biotechnology information) at NIH (National institute of health) means to collect the maps of epigenetic modification and the occurrence across the human genome.

>AEQ92754.1 sedoheptulose-1, 7-bisphosphatase, partial [*Achillea Millefolium*]

IYGPRTTYVIAIKGFPGTHEFLLLDEGKWQHVKETTEISEGKMFSPGNLRATFDNPDYEKLINYYVKEKYT

LR

(Sources—FASTA sequence related to *Solanum lycopersicum* retrieved from NCBI.)

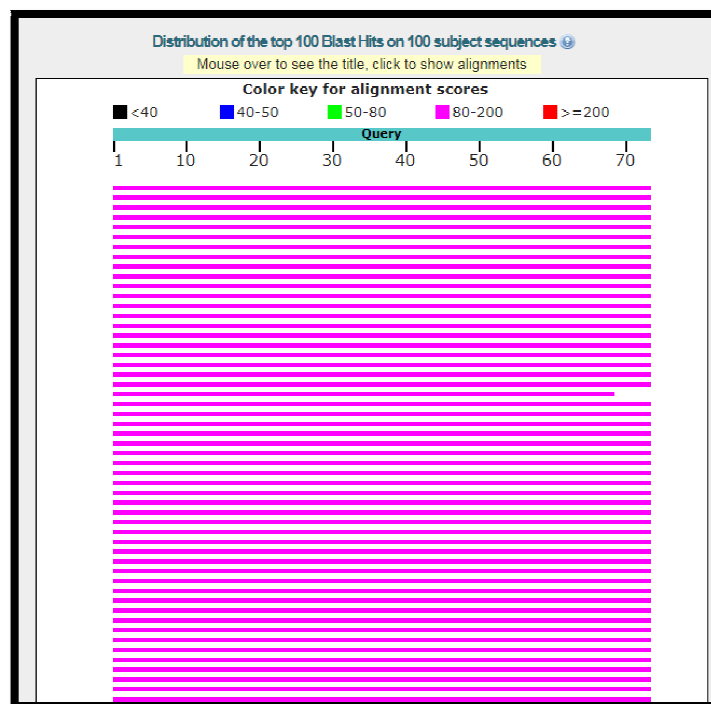


Figure 1: Graphical Summary of BLAST

Advanced Features of Sequence Analysis

The features table block is an important section in a nucleotide and protein sequence entry. The alignment would exist in the database by chance, if the database 610 sequence, then might expect that alignment occur may be 7 times. Score of the protein sequence, measure of similarity between the sequences. It is a statically calculation based on the quality of alignment obtained from one database and prediction sequence. An e-value of $1e-3$ is saying that there is a 0.001 chance that that alignment would exist in the database by chance, that is, if the database contains 10000 sequences. The expectation value of the *Achillea* protein sequence $5e-47$ annotate that there statical calculation based on the quality of the alignment.

CONCLUSIONS

The protein sequence of *Achillea* calculate the maximum number of aligned biological database. The expectation value of *Achillea* sequence increased from default value and size of the database showing as graphical form pink bars. *Achillea* sequence closer the E-value is towards 0, the better the alignment. It is a similarity search program for comparing biological sequence such as protein sequence.

REFERENCES

1. Acevedo-Rodriguez P., Strong M.T. 2012. *Catalogue of the Seed Plants of the West Indies*. Smithsonian Contributions to Botany, 98:1192 pp. Washington DC, USA: Smithsonian Institution. <http://botany.si.edu/Antilles/WestIndies/catalog.htm>
2. Akram, M., 2013, Minireview on *Achillea Millefolium* Lin. *Journal of Membrane Biology*, 246:661-663
3. Baxevanis, D. Andreas, *Quellete Fracis B.F.*, A Practical guide to the Analysis of gene and Proteins. 3rd Edition October 2004, Published by Wiley, John and Sons.

4. Alou, I. N., Van Asten, P. J., & Tenywa, M. M. (2014). Biophysical and crop management gradients limiting yields of east African highland banana (*Musa Spp. Aaa-Ea*) within farms in low input cropping systems. *International Journal of Agricultural Science and Research (IJASR)*, 4(3), 27-43.
5. Casey, R. M. (2005). "BLAST Sequences Aid in Genomics and Proteomics". *Business Intelligence Network*.
6. Higgins Des, Taylor Willie, 2000, *Bioinformatics: Sequence structure and database practical approach* " 1st Edition October 2000, Published by Oxford University Press.
7. Jean-Michel Claverie & Cedric Notre dame, *Bioinformatics for Dummies*, Wiley Publishing
8. Lipmann, D.J., Pearson, WR (1985). "Rapid and sensitive protein similarity searches". *Science* 227 (4693): 1435–. doi:10.1126/science.2983426. PMID 2983426.
9. MountDavid 2004, *Bioinformatics:-sequence & Genome Analysis*", published by Cold spring Harbour laboratory press.
10. Ara, T., & Choudhary, A. An attempt to propagate and conserve with *Ania Somnifera* (l). *Dunal: a polypharmaceutical plant through in vitro culture*.
11. Pearson, W.R. Lipman, D.J. (1988). "Improved tools for biological sequence comparisons". *Proceedings of the National Academy of Sciences of the United States of America* 85 (8): 24448. doi:10.1073/pnas.85.8.2444. PMC. PMID 3162770.
12. Kumari Uma 2018, "Quality impact studies of Variability on Genome Sequence of Tomato" Department of Biotechnology, JRU Jharkhand.
13. Pearson, W.R., Lipman, DJ (1988). "Improved tools for biological sequence comparisons". *Proceedings of the National Academy of Sciences of the United States of America* 85 (8): 2444-8. doi: 10.1073/pnas.85.8.2444. 280013. PMID 3162770.
14. Rick, C.M., Yader J.T., *Annu Rev Genet* 1988 *Classical and Molecular Genetics of tomato*.
15. Parameswari, M. *Textile and dye industry effluent, sludge and amendments on dehydrogenase and phosphatase activity of soil under sunflower crop*.
16. Taylor Willi and D.S. Higgins, 2000, *Bioinformatics :Sequence structure and database practical approach* "1st Edition October 2000, Published by Oxford university press.